Linguistic Indicators of Minor Depression in Chinese Older Adults' Life Stories

CHE Yiran
School of Foreign Studies, Tongji University, China

The University of Osaka - Tongji University
Joint International and Interdisciplinary Forum:
Toward the Integration of Society and Knowledge
Official Forum Proceedings

Abstract

Background: Minor depression is common in older adults, but often undetected, and linked to adverse outcomes. Language analysis can reveal cognitive and affective processes underlying mood disorders, and life stories provide rich data for detection.

Methods: We collected life story narratives (High, Low, and Turning points) from 34 healthy and 16 depressed Chinese-speaking older adults. Lexical, semantic, and sentiment features were analyzed using mixed ANOVAs with Bonferroni correction. Five machine learning algorithms then identified optimal feature combinations, evaluated by AUC, accuracy, and F1-score.

Results: Depressed participants used more modifiers in High and Low points and had lower mean dependency distances in Turning points. Low-point stories contained more utterances and higher sad intensity. Machine learning models performed best with Low and Turning point features; an XGBoost model with 7 features achieved highest performance in Turning-point narratives (accuracy 0.840, F1 0.743, AUC 0.829).

Conclusion: NLP-based models show promise for detecting minor depression in older adults via life stories. Future work should validate findings in larger, diverse cohorts.

Keywords: Life story, Older adult, Depression, Natural language processing

Introduction

By the year 2050, according to the United Nations (UN), one in six persons will be 65+ years of age. Given this increasing number of people entering the worldwide aging community, coupled with lower birth rates, there is concern about the significantly increasing old-age dependency ratio, especially in China, whose population aged 60 and above currently has surpassed 264 million, accounting for over 18% of the total population. One in four older adults experiences issues with their mental health, contributing to 23% of the total global burden of disease related to older adults (Gharib et al., 2024). Mental health problems significantly impact older adults' ability to carry out basic daily living activities, reducing their independence, autonomy, and quality of life. The first step to mitigate these negative consequences is to make a diagnosis. Unfortunately, mental health problems are often undiagnosed and untreated, leaving many older people without proper help (Rashedi et al., 2018).

Words convey important information about people's thoughts and feelings, and can reflect the way in which people experience the world. Online posts and oral speech about mental states, worries, or life stories have been employed to identify mental illness. Scholars have found distinguishing acoustic (e.g., decrease in speaking rate, lower vocal intensity range, greater jitter) and linguistic characteristics (e.g., choice of words, sentiment intensity) of patients with various mental disorders (Cummins et al., 2015; Low et al., 2020; Cummins et al., 2023). Recently, the development of natural language processing (NLP) tools has greatly facilitated the process of automatic feature extraction from speech data (Martínez-Nicolás et al., 2021), and machine learning techniques offer a potent tool for disease detection, making the analysis of language features more convenient.

Minor depression, characterized by depressive symptoms that do not fulfill the severity requirement for major depression, is the most common type of depression experienced by older adults (Beekman et al., 1999; Brenes et al., 2007). Minor depression is a risk factor for major depression, and is associated with an increased number of chronic diseases, functional limitations, increased severity of medical illness, and physical disability (Langa et al., 2004; Brenes et al., 2007). This study harnesses the NLP and machine learning tools to advance research on understanding and distinguishing minor depression among older adults through connected speech, with the goal of improving the early detection of depression and care for the growing aging population.

Literature Review

Researchers have widely studied linguistic style in relation to mental disorder by leveraging the development of NLP tools such as Linguistic Inquiry and Word Count (LIWC) and Python code.

The heightened self-focus theory of Pyszczynski and Greenberg (1987) suggested that people who are depressed tend to perseverate about themselves. Thus, the use of personal pronouns is suggested to connect with affective states. First-person singular pronouns has been identified as a consistent "hallmark of depression" across various studies (Newell et al., 2017; Viduani et al., 2025). However, some studies found null or opposite effects on their correlation (e.g., Rodriguez et al., 2010; Trifu et al., 2024). Tackman et al. (2019) found that the depression-first-person pronoun effect was substantially reduced when controlled for negative emotionality, which suggested that self-referential language using first-person singular pronouns may be better construed as a linguistic marker of general distress proneness or negative emotionality rather than as a specific marker of depression. On the contrary, first-person plural pronoun usage is linked with better perceived support and an expanded sense of self (Aron et al., 2004; Simmons et al., 2005).

Sentiment features, such as sentiment dominance, sentiment valence, and the use of positive or negative words, are one of the most sensitive linguistic signs of depressive mood (Gumus et al., 2023). Increased negative and positive words may suggest greater emotional awareness, i.e. the ability to identify and communicate one's emotions (Kranzler et al., 2016). Emotional blunting, a common feature in depression (Christensen et al., 2022), may also help explain why a decrement in emotional words was associated with higher severity of depressive symptoms (Weintraub et al., 2023). Depressed individuals were shown to use fewer positive or emotional words (Rude et al., 2004), and this finding is recently been verified in a non-clinical, normative population (Rodriguez et al., 2010; König et al., 2022; Gumus et al., 2023). Improved depressive states have been found to be associated with heightened use of positive emotion words and little use of negative emotion words (Himmelstein et al., 2018; Hartnagel et al., 2025). A study, however, suggested a greater use of both positive and negative emotion words with improvements in depressive symptoms (Weintraub et al., 2023), but this divergent finding in negative emotion may result from the limitation of the LIWC method, whose analysis is independent of context.

Studies also revealed that lexical richness, ratio of content words, and the organization of text decreased linearly across the severity of mental disorders, and depressiveness correlated with average sentence length and coefficient of coherence (Litvinova et al., 2016; Smirnova et al., 2018; Havigerová et al., 2019; Du, 2023). Barnes et al. (2007) found that before committing suicide, a person used much longer sentences and more question markers in his letters. Similarly, Pestian et al. (2010) and Kim et al. (2019) found that suicide notes had more words or phrases per sentence, fewer complete sentences, and more simple and common words per sentence.

Compared with lexical features, the psychological relevance of syntactic structures in language has not been broadly addressed. Syntactic decisions that participants make in the sequential formulation of clauses can indicate the extent to which they integrate past events in the narrative reconstruction of experiences, and thus the causal understanding reflected in syntax may demonstrate narrators' ability in rationalization and internalization. Zinken et al. (2010) used brief free texts written about current and previous problems to predict the likelihood of completion in guided self-help. The authors concluded that participants using more non-finite complement clauses were more likely to complete the program, and among those who completed, the use of causation words and complex syntax (e.g., adverbial clauses) predicted improvement. Schneider et al. (2023) found that patients with schizophrenia spectrum disorder produced significantly less complex sentences and significantly fewer types of complex sentences compared to healthy controls and major depression patients, but no significant difference was found between healthy controls and depressive patients.

From the perspective of neuropsychology, experimental studies have found that depressed individuals are most prominently affected in the domains of executive function, memory, and processing speed (Potter & Steffens, 2007), results that are consistent with meta-analyses of the broader literature (Feil et al., 2003). Complex sentences require higher cognitive competence. Dependency distance refers to the linear distance measured in terms of the number of intervening words between a parent (governor) and a dependent (Liu, 2008). Mean dependency distance (MDD) is the absolute value sum of the dependency distances divided by the number of dependency relations and is already found to be correlated with aging-related cognitive decline. Together with other linguistic features, MDD can successfully identify Chinese older adults with mild cognitive impairment (Yih et al., 2025). Thus, it can be hypothesized that affected by cognitive decline, depressive individuals may produce more simplified dependency structures when compared with healthy controls.

Based on linguistic markers, supervised learning-based algorithms like Support Vector Machine (SVM), Random Forest, and Logistic Regression, and text mining approaches were able to achieve good results in differentiating depressed and healthy individuals. Most of the studies conducted their experiments based on lexical, semantic, and sentiment features related to depression collected from online forums. Wang et al (2013) focused on 468 users of the Sina Microblog, employing three types of approaches: rules, trees, and Bayes, all of which achieved accuracies around 80%. They discovered that the times of mentioning others was highly predictive of depression. Nguyen et al. (2016) found that when LIWC and topics were used as feature sets in blog post classification, performance was effective, with accuracies of 88% and 93%, respectively. Choudhury et al. (2013) conducted depression detection using SVM and obtained good performance, with an accuracy of 70%. As an overall evaluation, computational models using supervised learning classifiers were able to achieve a 70-90% range accuracy (for more literature, see Arachchige et al., 2021 and Liu et al., 2022).

Previous studies gave us a relatively clear conclusion that a person's mental state is correlated with their language production in various linguistic domains. However, most studies were conducted on adolescents and online forums, whereas the linguistic features of older adults with different levels of depression are rarely studied.

Moreover, although language data collected from online forums, social media, clinical interviews, or brief talks like reporting current feelings via Smart phone (Hartnagel et al., 2024) or Speech Apps (Gumus et al., 2023) showed its effectiveness in identifying mental disorder, there's a key limitation of such data: it is often generated after individuals have already become aware of their psychological struggles or been diagnosed as patients of depression. Unlike spontaneous online posts or clinical interviews, life narratives can be elicited from individuals in daily settings, which allows us to detect the underlying language patterns of depression before a person self-identifies as having a problem. This shifts the application of language analysis from confirming known conditions to a proactive screening mechanism, potentially facilitating earlier support and intervention for at-risk individuals.

Also, linguistic features are undoubtedly related to the language used. For example, Trifu et al. (2024) found no significant differences in the use of the first-person singular personal pronouns between depressive patients and controls, and the authors attributed this result to the specificities of the Romanian language. Romanian language uses mostly non-accentuated or incomplete forms of first-person singular pronouns and uses self-reporting via inflexional and morphological verbal forms included in the conjugation, and morphemes, respectively. The subject of the sentence is often implicit in the Romanian language; therefore, the actual use of a personal pronoun with the verb is not necessary. Similarly, Lyu et al. (2023) found that the first-person singular pronouns contribute to depression detection in Chinese social media. Reasons for this result may be that the absence of a subject was very common in Chinese social media, and first-person singular pronouns in Chinese are often used as pragmatic empathetic deixis to narrow the psychological distance between addresser and addressee, but more data is needed to confirm this finding.

The present study addresses research gaps by investigating how Chinese older adults with and without depression narrate their life stories. This study comprehensively compares lexical, syntactic, and sentiment differences of older adults with and without minor depression, and discusses the effect of different kinds of life stories on older adults' language. Also, this study tested the possibility of using machine learning algorithms to detect minor depression using life stories.

Methodology

Participants

Fifty-four Chinese-speaking older adults were recruited online or from local communities in China. Each participant got 30-50 yuan or household goods as honorarium. Four participants were excluded from the data set due to their incompleteness of narrative tasks or poor recording quality, leaving 50 older adults as participants in this study. Participants' mean age was 67.36 (from 60 to 93 years old), with 10.15 mean years of education (from 3 to 17 years).

Depression was tested by the 15-item geriatric depression scale (GDS-15), which is a useful tool for multi-category classification of geriatric depression as normal, minor depressive disorder (MnDD), and major depressive disorder (MDD) in communities. Participants with GDS<5 were grouped as healthy in this study; participants with GDS≤10 were grouped as minor depressive older adults, resulting in 34 participants grouped as healthy and 16 grouped as minor depressed. The mean depression score was 3.6 among all participants (from 1 to 10). Detailed statistics are shown in Table 1.

Group	Healthy (n=34)	Depressed (n=16)	t /χ^2	*p*
Age (year)	66.97±7.06	68.19±9.28	-0.513	0.610
Education (year)	10.13±3.36	10.19±3.54	-0.053	0.947
Gender (female/male)	24/10	11/5	0.018	0.895
GDS - 15	2.29±0.97	6.38±1.86	-8.273	<0.001

Table 1: Demographic information of the participants.

Interview

This research used the first three prompts from the part of McAdam's (2008) Life Story Interview ("Key Scenes in the Life Story"). This is a revised and structured procedure in which the interviewer asked a series of questions designed to probe for the most important episodes in a person's life story. A general prompt was offered to participants and they were encouraged to narrate without interruption. The interviewer inquired about (a) the most positive experience the participant had had so far (High-point), (b) the most negative or upsetting experience (Low-point), and (c) the experience that marked the biggest change in the participant's life so far (Turning-point). The exact prompts closely followed McAdams' (2008) interview.

Interview recordings were transcribed automatically and manually corrected. Transcriptions were then segmented into three files according to High-point, Low-point, and Turing-point tasks. The interviewer's speech and participants' filler phrases were removed from the texts before parsing. These transcription segments were then processed using Stanza (version 2.0) package, a Python-based NLP tool, for sentence segmentation, tokenization, part-of-speech tagging, dependency parsing, etc (Qi et al., 2020). Finally, automatic linguistic unit identification and feature extractions were done by self-compiled Python codes.

Linguistic Features

Based on previous research's findings, we extracted 21 linguistic features in the present research, which can be divided into three classes: lexical, syntactic, and sentiment. Detailed descriptions of linguistic features are shown in Table 2.

Domains	Subdomains	Features (codes)	Calculation methods
Lexical	Lexical diversity	MATTR-50	$\sum_{i=1}^{N-n+1} \frac{V_i}{n \times (N-n+1)}$ n (the window size) is 50 and V_i is the number of types in the i^{th} window.
	POS features	the ratio of nouns (noun%)	The number of nouns/ the number of total words
		the ratio of verbs (verb%)	The number of verbs/ the number of total words
		the ratio of adjectives (adj%)	The number of adjectives/ the number of total words
		the ratio of adverbs (adv%)	The number of adverbs/ the number of total words
		the ratio of conjunctions (conj%)	The number of conjunctions/ the number of total words
		the ratio of pronouns (pron%)	The number of pronouns/ the number of total words
		the ratio of first-person singular pronouns (I%)	The number of first-person singulars/ the number of pronouns
		the ratio of first-person plural pronouns (we%)	The number of first-person plural/ the number of pronouns
		open-closed words ratio (open-closed%)	the number of open-class words/ the number of closed-class words
		the ratio of temporal words (time%)	the number of temporal words/ the number of total words

Syntactic	Sentence	number of utterance (U)	The total number of utterances
		mean length of utterances (MLU)	The mean number of words in utterances
	Dependencies	mean dependency distance of discourse (MDD)	$\frac{1}{n-s}\sum_{i=1}^{n} \lvert DD_i \rvert$ n is the total number of words in the text, s is the total number of sentences in the text
		the ratio of adverbial modifiers (AdvMod%)	The number of adverbial modifier/ the total number of dependencies
		the ratio of adjectival modifiers (AdjMod%)	The number of adjectival modifier/ the total number of dependencies
		the ratio of nominal modifiers (NomMod%)	The number of nominal modifier/ the total number of dependencies
		the ratio of head-final dependencies (head-final%)	The number of head-final dependencies/ the total number of dependencies
Sentiment	Intensity	negative intensity (negative)	
		good intensity (good)	
		sad intensity (sad)	

Table 2: Linguistic features extracted from transcripts

Statistical Analyses

Univariate 2×3 mixed ANOVAs (one per measure of interest) were conducted, which included groups (Healthy, Depressed) and tasks (High-point, Low-point, Turning-point), and post-hoc comparisons using Bonferroni corrections.

Machine learning methods combined with feature selection algorithms were employed to identify optimal composite features and feature combinations. Prior to modeling, z-score normalization was performed to enhance training stability and model performance. Five classifiers were implemented: Linear Discriminant Analysis (LDA), Logistic Regression (LR), Random Forest (RF), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost).

Class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic minority class samples through interpolation between existing instances. The k-neighbors parameter in SMOTE was dynamically adjusted based on the minority class sample size in each training fold to prevent over-sampling artifacts. To ensure robust performance estimation, we implemented 5-fold Stratified Cross-Validation, preserving class distribution in each split.

For feature selection, Recursive Feature Elimination (RFE) was applied to the complete feature set to systematically evaluate model performance across different feature subset sizes and identify the optimal feature combination. RFE operates by recursively eliminating the least important features based on classifier-derived importance rankings, with comprehensive evaluation performed across feature subsets ranging from 1 to 21 features for each algorithm.

Model performance was comprehensively assessed using accuracy, precision, recall, F1-score, and AUC metrics. Hyper-parameter optimization was conducted for each classifier using grid search with cross-validation. All pre-processing steps, including feature selection and SMOTE resampling, were strictly applied within cross-validation loops to prevent data leakage and ensure unbiased performance estimation.

Results

Differences in groups and tasks

Univariate ANOVAs were conducted on 21 linguistic features on group (1: Healthy, 2: Depressed) and task (1: High-point, 2: Low-point, 3: Turning-point) respectively. The results showed no significant interaction effect on group and task ($p>0.05$).

Mild differences in life narrative production were found between healthy and depressed older adults. Notably, in the High-point life stories, the ratio of adjectives was significantly higher in the depressed group than in the healthy group ($F(1,48)=5.318$, $p=0.023$, $\eta^2=0.036$). In the Low-point task, the ratio of adverbial modifiers was also significantly higher in the depressed group ($F(1,48)=5.867$, $p=0.017$, $\eta^2=0.039$). In the

Turning-point task, MDD was significantly higher in the healthy group than in the depressed group ($F(1,48)=5.027$, $p=0.026$, $\eta^2=0.034$).

There were significant differences in syntactic and sentiment indicators between different tasks within the Healthy group. The number of utterances ($F(2,96)=4.969$, $p=0.008$, $\eta^2=0.065$), and sad intensity ($F(2,96)=9.862$, $p<0.001$, $\eta^2=0.120$) were significantly higher in the Low-point Task than in the High-point and Turning Point tasks. MDD ($F(2,96)=2.415$, $p=0.092$, $\eta^2=0.032$) was higher in the Turing-point Task and in the Low-point Task. Negative intensity was higher in the Low-point Task than in the Turning-point Task ($F(2,96)=2.721$, $p=0.069$, $\eta^2=0.036$).

In the depressed group, the ratio of adjectives was significantly higher in the High-point Task than in the Low-point Task ($F(2,96)=2.878$, $p=0.059$, $\eta^2=0.038$). Similar to the Healthy group, sad intensity ($F(2,96)=5.290$, $p=0.006$, $\eta^2=0.068$) was significantly higher in the Low-point Task than in the High-point and Turning Point tasks.

Machine learning classification

Many models and feature combinations exhibit distinct performances in terms of two evaluation metrics: accuracy and AUC. However, accuracy is a deceptive and unreliable metric for imbalanced data, whereas AUC measures the model's ability to distinguish between classes, independent of the class distribution or the classification threshold. A high F1-score indicates that the model achieves a good balance between finding most of the positive cases (Recall) and ensuring its positive predictions are reliable (Precision). Thus, after filtering the models based on the highest AUC, models with the highest F1-score are shown in Table 4.

In the High-point narrative task, the model with the highest AUC was achieved by an SVM model with 1 feature (negative intensity), reaching the AUC value of 0.80 (Accuracy: 0.70, Precision: 0.55, Recall: 0.82, F1: 0.65). For the Low-point narrative task, the highest AUC was achieved by an XGBoost model with 4 features (U, sad, pronoun%, AdvMod%), and its AUC reached 0.83 (Accuracy: 0.76, Precision: 0.70, Recall: 0.70, F1: 0.64). XGBoost models performed best in the Turning-point task. A model with 6 features reached the highest AUC, reaching the AUC value of 0.85, mean ACC reaches 0.80, Precision: 0.79, recall: 0.62, F1: 0.66, and the feature set includes MLU, negative, sad, time%, MDD, and NomMod%.

Task	Model	ACC	precision	recall	F1	AUC	features
High	XGBoost	0.84	0.77	0.73	0.74	0.74	3: MLU, adj%, verb%
Low	XGBoost	0.80	0.68	0.80	0.70	0.78	10: U, Good, Sad, MATTR-50, adv%, pron%, time%, verb%, MDD, AdvMod%
Turning	XGBoost	0.84	0.84	0.75	0.74	0.83	7: MLU, negative, Sad, time%, MDD, NomMod%, we%

Table 3: Feature sets selected in the three tasks

Discussion

When narrating High-point life stories, older adults with depression showed a higher ratio of adjectives than healthy older adults. It is believed that participants with depression recalled significantly more summary memories in response to a request for a positive self-defining memory than did participants with lower depression scores (Moffitt et al., 1994). This over-general memory, characterized by difficulty in retrieving specific events with more concrete details (Bucci &Freedman,1981), may cause the frequent use of adjectives instead of concrete verbs or nouns referring to what exactly happened (Yahya & Rahim, 2023). Also, positive memories in depressed individuals sometimes exhibit an interwoven with nostalgia, sadness, or a sense of loss for past happiness, creating a linguistic demand for diverse adjectives to depict complex psychological states. However, the negative intensity and sad intensity showed no significant difference.

In the Low-point task, the ratio of adverbial modifier dependencies (AvdMod%) was significantly higher in the depressed group than in the healthy group, and absolutist and causal thinking may be related to this result. Cognitive theory of depression proposes that absolutist thinking is one of the cognitive biases that depressed individuals tend to exhibit. It is characterized as polarized, all-or-nothing thinking with no middle ground for other possibilities, exhibiting difficulty tolerating ambiguity and ambivalence. Absolutist words like adverbial intensifiers and modal verbs were more prevalent in anxiety, depression, and suicidal ideation forums than in non-affective group forums, and are more reliable markers of depression than negative emotion words (Al-Mosaiwi &Johnstone, 2018).

Turning points are specific events that are perceived to alter the normal flow and direction of one's life. Compared with High- and Low-points, Turning-point narration may be more complex and requires more cognitive and psychological process, for it emphasizes meaning-making (Thorne & McLean, 2002; Throne et al., 2004), i.e., finding the significance of events with high tension and conflict to people's following lives or values. This study found that MDD was higher among healthy participants than depressed individuals in the Turing-point task. Higher MDD suggests that healthy older adults' stories are syntactically complicated, using more modifiers and conjunctions to integrate disparate events, causal relationships, and emotions into a coherent narrative framework. Different from a higher degree of causality among depressed participants in the Low-point task, which suggests their rumination as a response to distress, when there's a higher need for meaning-making, older adults with depression revealed a more frequent addressing of descriptive rather than analytic thought strategies.

Studies found that major depressive disorder among older adults is associated with reduced syntactic abilities (Xu et al., 2023). People with mild depression predominantly used single-clause sentences and reduced sentences, and compound sentences were more commonly used than complex ones. Atypical/inverse word-order usage was also revealed in depressive patients (Smirnova et al., 2018). Moreover, MDD has been shown to be a key textual measure reflecting cognitive abilities relating to memory (Liu, 2008) and is found to be linked with older adults' cognitive impairment (Yih et al., 2025). Systematic review and meta-analyses revealed that the severity of depression is related to reduced cognitive performance in episodic memory, executive function, and processing speed (McDermott & Ebmeier, 2009; Rock et al., 2014). Thus, significantly lower MDD in the depressed group may also result from the decline in several cognitive domains as a side-effect of depression.

Negative intensity and first-person pronouns are widely established as markers of depression in highly influential literature (e.g., Tausczik & Pennebaker, 2010; Al-Mosaiwi & Johnstone, 2018; Chung & Pennebaker, 2019). In this study, depressed older adults produced higher sad intensity and lower good intensity when compared with the healthy group, but the use of first-person pronouns was irregular, and all these data showed no significance. This divergence may stem from different task paradigms. Online self-narratives or self-generated discourse about mental state essentially elicit emotion-driven speech, yielding high emotional density, emphasizing the evaluative function of narrative. In contrast, semi-structured life review requires participants to give more referential information about their experiences. Participants need to recall, select, organize, and make sense of self-focused personal experiences and thus increase the focus on self but attenuate the focus on emotion, consequently leading to a dilution of emotional salience in the language.

When looking at the differences between tasks, language in the Low-point narrative task is significantly different from High- and Turing-point narratives, especially in terms of the number of utterances and sad intensity. The controversy of the characteristics of negative experiences has long been debated. Trauma Superiority theory holds that traumatic events are remembered better than non-traumatic events, and are assumed to be "superior" to other memories since they would be recalled more vividly and accurately (Sotgiu & Rusconi, 2014). By contrast, Trauma theory asserts that traumatic memories are more fragmented, disorganized, and less coherent compared to nontraumatic memories. Poorer memory for trauma has been linked to the activation of defense mechanisms such as, for example, dissociation and repression (Whitfield, 1995; Alpert et al., 1998), and studies have found that positive memories contain more details and are remembered more clearly and more vividly than traumatic or intensely negative memories (Tromp et al., 1995; D'Argembeau et al., 2003). In the present study, Low-point narratives have more utterances than speech in the other tasks, especially in the healthy group, but POS features and lexical diversity showed no significant difference, suggesting neither the superiority nor inferiority of Low-point experiences among older adults with minor depression.

Detection of minor depression using ML algorithms

As an overall evaluation, computational models developed for detecting depression using learning classifiers and text mining approaches were able to achieve a 70-90% range accuracy (Arachchige et al., 2021). Initializing RFE estimators, 5-fold stratified cross-validation, and SMOTE for handling class imbalance, the current study has obtained several models with certain feature combinations that reach a mean accuracy and AUC higher than 0.80. However, due to the imbalanced data set, we should be cautious of the mean accuracy of models and put more emphasis on the AUC and F1 value. Our best result is in an XGBoost model in Turning-point narratives with 7 features, reaching an AUC value of 0.8286 and F1 value of 0.7433, but the model's performance, potentially unstable due to the small sample size, is lower than established benchmarks. For example, Khan & Alqahtani (2024) reported near-perfect metrics (Accuracy: 0.994, F1: 0.991) using TF-IDF with linear models, and Fatima et al. (2019) achieved an F1 of 0.9171 with an MLP classifier. This gap can be largely attributed to fundamental differences in data provenance. Unlike previous studies that leveraged explicitly disease-centric discourse from platforms like Facebook, Twitter, Reddit, or online support forums (OSF), our data comprises daily life narratives from a minor depression cohort. In such narratives, depressive linguistic manifestations are inherently more indirect and nuanced, posing a greater challenge for classification.

Linguistic feature sets of the best-performing models in different tasks (Table 5) are inconsistent, whereas different tasks relied on distinct feature types. The High-point task was primarily driven by syntactic and lexical features (MLU and POS features). Models with the best performance in the Low-point and Turning-point tasks depended on linguistic dimensions combining sentiment, lexical, and syntactic features. For model and task selection, XGBoost consistently delivered top performance across all tasks, and models in the Low- and Turing point tasks performed better than in the High-point task.

Limitations

This study has several limitations that should be considered when interpreting the results. Firstly, the generalizability of the findings is constrained by the relatively small sample size, particularly the group of older adults with minor depression (n=16). Although statistical corrections and machine learning techniques to address class imbalance were employed, the limited number of participants may affect the stability and accuracy of the observed effects and the machine learning classification performance.

Secondly, the selection of linguistic indicators was not exhaustive. Certain features salient in depression research, such as the use of absolutist words (Al-Mosaiwi & Johnstone, 2018), concrete and abstract words (Tausczik &Pennebaker, 2010), or coherence and redundancy (Smirnova et al., 2018), were not examined due to the current lack of reliable, validated Chinese word lists for such categories. Subclinical depression is also found to be related to a high use of words that indicate cognitive processing and meaning-making concerns (e.g., because, realize, understand) and attributions about negative events found in life-narrative stories (Adler et al., 2006).

Finally, the analysis primarily focused on how participants narrated their life stories rather than what they said. Future research would benefit from a larger sample size, a wider array of linguistic features, and the incorporation of content analysis methods, such as word frequency statistics followed by qualitative thematic analysis. This integrated approach would more fully illuminate the core themes in older adults' life narratives and their underlying psychological states.

Conclusion

Depression in older adults is a serious health problem with a poor prognosis. Life review shows the potential of reducing depressive symptoms, and improving life satisfaction in older adults through the course of looking for meaning (Pot et al., 2010; Jiang et al., 2024), but life narrative does more than that. This study found that Chinese older adults with minor depression used more modifiers in their High- and Low-point narratives, but produced simplified dependency structure in Turning-point narratives. Machine learning models, especially the XGBoost, can accurately detect the signs of depression using life narratives, further proving the significance of life review and demonstrating the possibility of using life stories in the assessment and prediction of minor depression among older adults.

References

Adler, J. M., Kissel, E. C., & McAdams, D. P. (2006). Emerging from the CAVE: Attributional style and the narrative study of identity in midlife adults. *Cognitive therapy and research*, 30(1), 39-51.

Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical psychological science*, 6(4), 529-542.

Alpert, J.L., Brown, L.S., & Courtois, C.A. (1998). Symptomatic clients and memories of childhood abuse: What the trauma and child sexual abuse literature tell us. *Psychology, Public Policy, and Law*, 4, 941–995.

Aron, A., McLaughlin-Volpe, T., Mashek, D., Lewandowski, G., Wright, S. C., & Aron, E. N. (2004). Including others in the self. *European review of social psychology*, 15(1), 101-132.

Barnes, D. H., Lawal-Solarin, F. W., & Lester, D. (2007). Letters from a suicide. *DeathStudies*, 31(7), 671–678.

Beekman ATF, Copeland JRM, Prince MJ. Review of community prevalence of depression in later life. *British Journal of Psychiatry*. 1999;174:307–311.

Brenes, G. A., Williamson, J. D., Messier, S. P., Rejeski, W. J., Pahor, M., Ip, E., & Penninx, B. W. (2007). Treatment of minor depression in older adults: a pilot study comparing sertraline and exercise. *Aging and Mental Health*, 11(1), 61-68.

Bucci, W., & Freedman, N. (1981). The language of depression. *Bulletin of the Menninger Clinic*, 45(4), 334.

Christensen, M. C., Ren, H., & Fagiolini, A. (2022). Emotional blunting in patients with depression. *Annals of General Psychiatry*, 21(1), 10.

Chung, C. K., & Pennebaker, J. W. (2019). Textual analysis. In H. Blanton, J. M. LaCroix, & G. D. Webster (Eds.), *Measurement in Social Psychology* (pp.153–173). Austin:Taylor &Francis Group.

Cummins, N., Dineley, J., Conde, P., Matcham, F., Siddi, S., Lamers, F., ... & RADAR-CNS Consortium. (2023). Multilingual markers of depression in remotely collected speech samples: A preliminary analysis. *Journal of affective disorders*, 341, 128-136.

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech communication*, 71, 10-49.

D'Argembeau, A., Comblain, C., & Van der Linden, M. (2003). Phenomenal characteristics of autobiographical memories for positive, negative, and neutral events. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 17(3), 281-294.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. In *Proceedings of the international AAAI conference on web and social media* (Vol. 7, No. 1, pp. 128-137).

Du, X. (2023). Lexical features and psychological states: A quantitative linguistic approach. *Journal of Quantitative Linguistics*, 30(3-4), 257-279.

Fatima, I., Abbasi, B. U. D., Khan, S., Al-Saeed, M., Ahmad, H. F., & Mumtaz, R. (2019). Prediction of postpartum depression using machine learning techniques from social media text. *Expert Systems*, 36(4), e12409.

Feil, D., Razani, J., Boone, K., & Lesser, I. (2003). Apathy and cognitive performance in older adults with depression. *International journal of geriatric psychiatry*, 18(6), 479-485.

Gharib, M., Borhaninejad, V., & Rashedi, V. (2024). Mental health challenges among older adults. *Adv Med Psychol Public Health*, 1(3), 106-107.

Gumus, M., DeSouza, D. D., Xu, M., Fidalgo, C., Simpson, W., & Robin, J. (2023). Evaluating the utility of daily speech assessments for monitoring depression symptoms. *Digital health*, 9, 20552076231180523.

Hartnagel, L. M., Ebner-Priemer, U. W., Foo, J. C., Streit, F., Witt, S. H., Frank, J., ... & Sirignano, L. (2025). Linguistic style as a digital marker for depression severity: An ambulatory assessment pilot study in patients with depressive disorder undergoing sleep deprivation therapy. *Acta Psychiatrica Scandinavica,* 151(3), 348-357.

Hartnagel, L. M., Emden, D., Foo, J. C., Streit, F., Witt, S. H., Frank, J., ... & Sirignano, L. (2024). Momentary Depression Severity Prediction in Patients With Acute Depression Who Undergo Sleep Deprivation Therapy: Speech-Based Machine Learning Approach. *JMIR Mental Health*, 11, e64578.

Havigerová, J. M., Haviger, J., Kučera, D., & Hoffmannová, P. (2019). Text-based detection of the risk of depression. *Frontiers in psychology*, 10, 513.

Himmelstein, P., Barb, S., Finlayson, M. A., & Young, K. D. (2018). Linguistic analysis of the autobiographical memories of individuals with major depressive disorder. *PloS one*, 13(11), e0207814.

Jiang, V., Galin, A., & Lea, X. (2024). Life review for older adults: an integrative review. *Psychogeriatrics*, 24(6), 1402-1417.

Khan, S., & Alqahtani, S. (2024). Hybrid machine learning models to detect signs of depression. *Multimedia Tools and Applications,* 83(13), 38819-38837.

Kim, K., Choi, S., Lee, J., & Sea, J. (2019). Differences in linguistic and psychological characteristics between suicide notes and diaries. *The Journal of General Psychology*, 146(4), 391–416.

König, A., Tröger, J., Mallick, E., Mina, M., Linz, N., Wagnon, C., ... & Peter, J. (2022). Detecting subtle signs of depression with automated speech analysis in a non-clinical sample. *BMC psychiatry*, 22(1), 830.

Kranzler, A., Young, J. F., Hankin, B. L., Abela, J. R., Elias, M. J., & Selby, E. A. (2016). Emotional awareness: A transdiagnostic predictor of depression and anxiety for children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, 45(3), 262-269.

Langa KM, Valenstein MA, Fendrick AM, Kabeto MU, Vijan S. Extent and cost of informal caregiving for older Americans with symptoms of depression. *American Journal of Psychiatry*. 2004;161:857–863.

Litvinova, T., Zagorovskaya, O., Litvinova, O., & Seredin, P. (2016). Profiling a set ofpersonality traits of a text's author: A corpus-based approach. In A. Ronzhin,R. Potapova, & G. Nemeth (Eds.), *Speech and Computer: Proceedings of the 18thInternational Conference*, SPECOM 2016 (pp. 555–562). Budapest, Hungary.Springer International Publishing.

Liu, D., Feng, X. L., Ahmed, F., Shahid, M., & Guo, J. (2022). Detecting and measuring depression on social media using a machine learning approach: systematic review. *JMIR Mental Health*, 9(3), e27244.

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2), 159–191.

Low, D. M., Bentley, K. H., & Ghosh, S. S. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope investigative otolaryngology*, 5(1), 96-116.

Lyu, S., Ren, X., Du, Y., & Zhao, N. (2023). Detecting depression of Chinese microblog users via text analysis: Combining Linguistic Inquiry Word Count (LIWC) with culture and suicide related lexicons. *Frontiers in psychiatry*, 14, 1121583.

Martínez-Nicolás, I., Llorente, T. E., Martínez-Sánchez, F., & Meilán, J. J. G. (2021). Ten years of research on automatic voice and speech analysis of people with Alzheimer's disease and mild cognitive impairment: a systematic review article. *Frontiers in Psychology*, 12, 620251.

McAdams, D. P. (2008, April). *The life story interview*.

McDermott, L. M., & Ebmeier, K. P. (2009). A meta-analysis of depression severity and cognitive function. *Journal of affective disorders*, 119(1-3), 1-8.

Moffitt, K. H., Singer, J. A., Nelligan, D. W., Carlson, M. A., & Vyse, S. A. (1994). Depression and memory narrative type. *Journal of Abnormal Psychology,* 103(3), 581.

Nanomi Arachchige, I. A., Sandanapitchai, P., & Weerasinghe, R. (2021). Investigating machine learning & natural language processing techniques applied for predicting depression disorder from online support forums: A systematic literature review. *Information*, 12(11), 444.

Newell, E. E., McCoy, S. K., Newman, M. L., Wellman, J. D., & Gardner, S. K. (2017). You sound so down: Capturing depressed affect through depressed language. *Journal of Language and Social Psychology*, 37(4), 451–474.

Nguyen, T., Venkatesh, S., & Phung, D. (2016, November). Textual cues for online depression in community and personal settings. In *International Conference on Advanced Data Mining and Applications* (pp. 19-34). Cham: Springer International Publishing.

Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., & Leenaars, A. (2010). Suicidenote classification using natural language processing: A content analysis.*Biomedical Informatics Insights*, 3, 19–28.

Pot, A. M., Bohlmeijer, E. T., Onrust, S., Melenhorst, A. S., Veerbeek, M., & De Vries, W. (2010). The impact of life review on depression in older adults: a randomized controlled trial. *International Psychogeriatrics*, 22(4), 572-581.

Potter, G. G., & Steffens, D. C. (2007). Contribution of depression to cognitive impairment and dementia in older adults. *The neurologist*, 13(3), 105-117.

Pyszczynski, T., & Greenberg, J. (1987). Self-regulatory perseveration and the depressive self-focusing style: a self-awareness theory of reactive depression. *Psychological bulletin,* 102(1), 122.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A python natural language processing toolkit for many human languages. *Proceedings of the 58th annual meeting of the association for computational linguistics: System demonstrations* (pp. 101–108). Association for Computational Linguistics.

Rashedi V, Malakouti SK, Rudnik A. (2018). Community mental health services for older adults in Iran. *J Kermanshah Univ Med Sci*, 22(2), e78947.

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological medicine*, 44(10), 2029-2040.

Rodriguez, A. J., Holleran, S. E., & Mehl, M. R. (2010). Reading between the lines: The lay assessment of subclinical depression from written self-descriptions. *Journal of personality*, 78(2), 575-598.

Rude, S., Gortner, E. M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8), 1121-1133.

Schneider, K., Leinweber, K., Jamalabadi, H., Teutenberg, L., Brosch, K., Pfarr, J. K., ... & Stein, F. (2023). Syntactic complexity and diversity of spontaneous speech production in schizophrenia spectrum and major depressive disorders. *Schizophrenia*, 9(1), 35.

Simmons, R. A., Gordon, P. C., & Chambless, D. L. (2005). Pronouns in marital interaction: What do "you" and "I" say about marital health?. *Psychological science*, 16(12), 932-936.

Smirnova, D., Cumming, P., Sloeva, E., Kuvshinova, N., Romanov, D., &Nosachev, G. (2018). Language patterns discriminate mild depression from nor-mal sadness and euthymic state. *Frontiers in Psychiatry*, 9, 105.

Sotgiu, I., & Rusconi, M. L. (2014). Why autobiographical memories for traumatic and emotional events might differ: Theoretical arguments and empirical evidence. *The Journal of Psychology*, 148(5), 523-547.

Tackman, A. M., Sbarra, D. A., Carey, A. L., Donnellan, M. B., Horn, A. B., Holtzman, N. S., ... & Mehl, M. R. (2019). Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of personality and social psychology*, 116(5), 817.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology,* 29(1), 24-54.

Thorne, A., & McLean, K. C. (2002). Gendered reminiscence practices andself-definition in late adolescence. *Sex Role*s, 46, 267–277.

Thorne, A., McLean, K. C., & Lawrence, A. M. (2004). When remember-ing is not enough: Reflecting on self-defining memories in late adoles-cence. *Journal of Personality*, 72, 513–542.

Trifu, R. N., Nemeș, B., Herta, D. C., Bodea-Hategan, C., Talaș, D. A., & Coman, H. (2024). Linguistic markers for major depressive disorder: a cross-sectional study using an automated procedure. *Frontiers in Psychology*, 15, 1355734.

Tromp, S., Koss, M.P., Figueredo, A.J., & Tharan, M. (1995). Are rape memories different? A comparison of rape, other unpleasant, and pleasant memories among employed women. *Journal of Traumatic Stress*, 8, 607–627.

Viduani, A., Cosenza, V., Fisher, H. L., Buchweitz, C., Mota, N., Piccin, J., ... & Kieling, C. (2025). First-person pronouns as linguistic markers of depression among Brazilian youths. *Brazilian Journal of Psychiatry*, 47, e20244058.

Wang, X., Zhang, C., & Sun, L. (2013, December). An improved model for depression detection in micro-blog social network. In 2013 IEEE 13th *international conference on data mining workshops* (pp. 80-87). IEEE.

Weintraub, M. J., Posta, F., Ichinose, M. C., Arevian, A. C., & Miklowitz, D. J. (2023). Word usage in spontaneous speech as a predictor of depressive symptoms among youth at high risk for mood disorders. *Journal of affective disorders*, 323, 675-678.

Whitfield, C.L. (1995). The forgotten difference: Ordinary memory versus traumatic memory. *Consciousness and Cognition*, 4, 88–94.

Xu, C., Wongpakaran, N., Wongpakaran, T., Siriwittayakorn, T., Wedding, D., & Varnado, P. (2023). Syntactic Errors in Older Adults with Depression. *Medicina*, 59(12), 2133.

Yahya, N. H., & Rahim, H. A. (2023). Linguistic markers of depression: insights from english-language tweets before and during the COVID-19 pandemic. *Language and Health*, 1(2), 36-50.

Yih, T., Yang, H., Huang, L., & Yao, Q. (2025). Identifying syntactic biomarkers of cognitive impairment in Mandarin-speaking older adults by applying machine learning approaches across multiple speech tasks. *Aphasiology*, 1-30.

Zinken, J., Zinken, K., Wilson, J. C., Butler, L., & Skinner, T. (2010). Analysis of syntax and word use to predict successful participation in guided self-help for anxiety and depression. *Psychiatry research*, 179(2), 181-186.

Contact: blake@tongji.edu.cn